

On a class of optimal transportation problems with infinitely many marginals*

Brendan Pass[†]

March 8, 2013

Abstract

We prove existence and uniqueness results for solutions to a class of optimal transportation problems with infinitely many marginals, supported on the real line. We also provide a characterization of the solution with an explicit formula. We then show that this result implies an infinite dimensional rearrangement inequality, and use it to derive upper bounds on solutions to parabolic PDE via the Feynman-Kac formula. This result can be used to provide model independent bounds on the prices of certain derivatives in mathematical finance. As an another application, we obtain refined phase-space bounds in quantum physics.

1 Introduction

In this paper we introduce and initiate the investigation of a novel problem - an optimal transportation problem of Monge Kantorovich type with infinitely many marginals. Given measures (or marginals) μ_t on \mathbb{R} , for all $t \in [0, 1]$, and a strictly convex function $h : \mathbb{R} \rightarrow \mathbb{R}$, we look for a stochastic process X_t with continuous sample paths, such that $\text{law} X_t = \mu_t$ for all $t \in [0, 1]$, which maximizes

$$E\left(h\left(\int_0^1 X_t dt\right)\right) \quad (MK_\infty)$$

Our motivation in studying (MK_∞) comes partially from an analogy with more familiar problems with finitely many marginals. The (two marginal) optimal transportation problem of Monge and Kantorovich is to couple a pair of probability measures together as efficiently as possible. This problem has many deep connections to a variety of areas of mathematics, such as geometry and PDE, and has found applications in many diverse fields, including economics, fluid mechanics, meteorology, and image processing [34][35]. Given two Borel

*The author was supported in part by a University of Alberta start-up grant.

[†]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1 pass@ualberta.ca.

probability measures μ_0 and μ_1 on smooth manifolds M_0 and M_1 , respectively, and a surplus function $b : M_0 \times M_1 \rightarrow \mathbb{R}$, the problem consists of finding the maximizer of the functional:

$$\int_{M_0 \times M_1} b(x_0, x_1) d\gamma \quad (MK_2)$$

among measures γ on $M_0 \times M_1$ which project to μ_0 and μ_1 under the canonical projections. Under mild conditions, a solution γ exists. Under somewhat stronger conditions, one can show that this solution is unique and is concentrated on the graph $\{x_1 = T(x_0)\}$ of a function over x_0 ; see Brenier [2], Gangbo and McCann [12], Gangbo [11], Caffarelli [4], McCann [23], and Levin [19]. In this case, we say that T solves Monge's optimal transportation problem. Let us note that the class of cost functions satisfying this condition includes $c(x_0, x_1) = h(x_0 + x_1)$, where h is strictly convex.

An interesting extension of (MK_2) is the problem of aligning several distributions with maximal efficiency. That is, given m Borel probability measures μ_i , for $i = 1, 2, \dots, m$ on smooth manifolds M_1, M_2, \dots, M_m , respectively, and a surplus function $b : M_1 \times M_2 \times \dots \times M_m \rightarrow \mathbb{R}$, maximizing

$$\int_{M_0 \times M_1 \times \dots \times M_m} b(x_1, x_2, \dots, x_m) d\gamma \quad (MK_m)$$

among measures γ on $M_1 \times M_2 \times \dots \times M_m$ which project to the μ_i under the canonical projections, for $i = 1, 2, \dots, m$.

For our purposes, it will be useful to have an equivalent formulation of this problem using more probabilistic language. From this perspective, we look for the $M_0 \times M_1 \times \dots \times M_m$ valued random variable (X_1, X_2, \dots, X_m) , such that $\text{law}(X_i) = \mu_i$ for all $i = 1, 2, \dots, m$, which maximizes the *expectation*:

$$E(c(X_1, X_2, \dots, X_m)) \quad (MK_m)$$

As in the $m = 2$ case, it is straightforward to prove the existence of a maximizer. The literature on (MK_m) , however, is much more limited than that on (MK_2) . The case that is most relevant to us here is when $M_i = \mathbb{R}^n$ for all $i = 1, 2, \dots, m$ and $c(x_1, x_2, \dots, x_m) = h(\sum_{i=1}^m x_i)$, for a strictly convex function h , this problem has been studied by Gangbo and Swiech [13] and Heinrich [16], using a duality result of Kellerer [17]; see also [24][18][30] and [31]. In this case, they proved that the optimal measure γ is unique and that it is concentrated on the graph $\{x_1, f_2(x_2), \dots, f_m(x_m)\}$ of a function over the first marginal, inducing what is often called a Monge solution. Since then, a handful of additional results have been obtained for other surplus functions, by Carlier [5], Carlier and Nazaret [6], and the present author [28][27].

It is natural to ask about the limit as the number of marginals, m , goes to infinity. Here, instead of prescribing m marginals, we prescribe a *continuum* of marginals μ_t on \mathbb{R}^n , for $t \in [0, 1]$, and rather than looking for a random variable (X_1, \dots, X_m) , we search for random *curves* X_t , (ie, stochastic processes) that satisfy $\text{law}(X_t) = \mu_t$ for each $t \in [0, 1]$. Heuristically, replacing the cost function

$h(\sum_{i=1}^m x_i)$ with $h(\int_0^1 x_t dt)$, we obtain problem (MK_∞) as the limit of (MK_m) as $m \rightarrow \infty$.

In this paper, we study (MK_∞) but restrict our attention to the case $n = 1$. The machinery developed here does not generalize easily to the higher dimensional case; however, in a subsequent paper, we address the higher dimensional case for the quadratic surplus (that is, $h(\cdot) = |\cdot|^2$), using very different techniques [25]. In the multi-marginal setting, when $n = 1$, Carlier [5] showed that, for a class of cost functions including strictly convex functions of the sum, the optimal maps $f_i : \mathbb{R} \rightarrow \mathbb{R}, i = 2, 3, \dots, m$ are all increasing. Our main result here (Theorem 3.1.1) is an infinite marginal analogue; there exists a unique optimal stochastic process X_t and it is concentrated on an increasing family of curves.

When $h(x) = e^x$, the reader may recognize expressions of form (MK_∞) from the Feynman-Kac representation of solutions to parabolic PDE (the Feynman-Kac expression is actually somewhat more complicated, but see the extension in subsection 3.2). Our main result implies a *stochastic rearrangement inequality* (Theorem 4.1.1); for any stochastic process, X_t , $E\left(h\left(\int_0^1 X_t dt\right)\right)$ can be bounded above by the same expression with the process X_t replaced by a particularly simple process. Therefore, this result can be used to derive upper bounds on the solutions to parabolic PDE which are straightforward to compute. Let us briefly mention that upper bounds on expressions of the form (MK_∞) can also be found by Jensen's inequality, but that our bound is always tighter than the Jensen's inequality bound. We discuss this in more detail in section 4.

Parabolic PDE arise frequently in various applied contexts, including mathematical finance and mathematical physics among many others. In mathematical finance, for example, the price of a zero coupon bond maturing in one year is given by $E[e^{-\int_0^1 X_t dt}]$; where the stochastic process X_t represents (stochastic) interest rate, given by some model. As another example, the value of an Asian call option on a stock maturing in one year is $E(h(\int_0^1 X_t dt))$, where X_t is the value of the stock price and $h(x) = \max(0, x - k)$ for some fixed k (the strike price) is piecewise linear. In some of these models, the expectation itself may be difficult to compute analytically, but the single time marginals can be readily determined and results in section 4 can be used to determine an upper bound on the interest rate simply by doing two 1-dimensional integrals.

Let us add that, from a financial perspective, the potential interest in these inequalities does not only lie in estimating integrals that are difficult to compute; it can also be used provide information about certain phenomena that is less model dependent than standard financial methods. Model independence in finance has recently seen quite a bit of attention. The goal here is to understand the difference between derivative prices arising from models with the same marginals, as is outlined in [1], [15] and [10]. Note that in these papers, while the authors restrict to discrete time models, they also impose the condition that the process X_t should be a martingale, which is natural in a financial context. Our optimal process, X_t for (MK_∞) , is not a martingale; nevertheless, it does provide an upper bound on the prices of appropriate derivatives arising from any model yielding the same marginals, and so may be relevant in their context.

Expressions of the form $E\left(h\left(\int_0^1 X_t dt\right)\right)$ also arise frequently in mathematical physics. In this context, upper bounds on these expressions can be used to derive phase space bounds in quantum mechanics. The best known example is perhaps Lieb's proof of the Cwickel-Lieb-Rosenbljum (CLR) bound, which itself is a special case of Lieb-Thirring inequalities, and can be used to derive the more general case [21]. The Lieb-Thirring inequalities are well known for their crucial role in Lieb and Thirring's famous proof of the stability of matter [22]. The general process for obtaining phase-space bounds is outlined nicely in the book by Simon [3]; very briefly, the key idea is to apply Jensen's inequality to expressions of the form (MK_∞) , then use Fubini's theorem to rearrange the upper bound into a desirable form. We show that our result here provides tighter bounds on these quantities.

In the next section, we outline our main assumptions and make our basic definitions. In section 3, we state and prove our main result. In the fourth section, we use our main result to derive a rearrangement inequality for stochastic processes, which can then be used to derive upper bounds on solutions to parabolic PDE via the Feynman-Kac formulas. In the final section, we describe a specific application of this inequality: new phase space bounds in quantum physics.

2 Preliminaries and assumptions

In this section, we outline our assumptions and preliminary definitions. Many of these assumptions can be relaxed, at the cost of more complicated proofs; in subsection 3.2, for example, we discuss two extensions of this theorem, including a relaxation of the first assumption.

Let Π be the set of all real valued, measurable stochastic processes X_t with continuous sample paths such that $\text{law} X_t = \mu_t$ for all $t \in [0, 1]$. We will assume:

1. The function $h : \mathbb{R} \rightarrow \mathbb{R}$ is strictly convex and C^2 smooth.
2. Each μ_t is absolutely continuous with respect to Lebesgue measure.
3. The support of each μ_t is a (possibly unbounded) interval: $\text{spt} \mu_t = [a_t, b_t]$, where a_t may be $-\infty$ and b_t may be ∞ .
4. The path of measures $t \mapsto \mu_t$ is weakly continuous.
5. The *first moments* and *h moments* of μ_t exist and are integrable in t . That is,

$$\int_0^1 \left| \int_{\mathbb{R}} x d\mu_t(x) \right| dt < \infty, \quad \int_0^1 \left| \int_{\mathbb{R}} h(x) d\mu_t(x) \right| dt < \infty.$$

Note that, for any process $X_t \in \Pi$, applying Jensen's inequality twice and using Fubini's theorem yields:

$$\begin{aligned}
h\left(\int_0^1 \int_{\mathbb{R}^n} x d\mu_t(x) dt\right) &= h\left(\int_0^1 E(X_t) dt\right) = h\left(E\left(\int_0^1 X_t dt\right)\right) \\
&\leq E\left(h\left(\int_0^1 X_t dt\right)\right) \\
&\leq E\left(\int_0^1 h(X_t) dt\right) = \int_0^1 E(h(X_t)) dt = \int_0^1 \int_{\mathbb{R}} h(x) d\mu_t(x) dt
\end{aligned}$$

Therefore, assumption 5 ensures that $E(h(\int_0^1 X_t dt))$ is finite for each competitor $X_t \in \Pi$.

Our last assumption requires a little more notation, which we develop now. For each $\lambda \in (0, 1)$, define the curve

$$x_\lambda(t) = F_t^{-1}(\lambda)$$

where F_t^{-1} is the inverse of the cumulative distribution function F_t of μ_t . Our final assumption can be interpreted as a regularity condition on the curve of measures μ_t , expressed in terms of the curves $x_\lambda(t)$:

6. The Riemann sums $\sum_{i=1}^N x_\lambda(\frac{i}{N})$ converge to $\int_0^1 x_\lambda(t) dt$ as $N \rightarrow \infty$, uniformly in λ on compact subsets of $(0, 1)$.

Note that weak continuity of the family μ_t (assumption 4) implies that the curve $t \rightarrow x_\lambda(t)$ is continuous and hence Riemann integrable. The Riemann sums above then converge to the integral, but the convergence is not necessarily uniform in λ .

Now, we define a particular stochastic process in Π , which, as we will prove in the next section, is actually optimal for (MK_∞) .

Our underlying probability space will be the interval $(0, 1)$, with Lebesgue measure. Define the process

$$X_t^{opt} = X_t^{opt}(\lambda) = x_\lambda(t), \text{ for } \lambda \in [0, 1]. \quad (1)$$

That is, the sample paths are the curves $t \mapsto x_\lambda(t)$, distributed uniformly in λ . Note that these sample paths are increasing in λ : whenever $\lambda \geq \bar{\lambda}$, $x_\lambda(t) \geq x_{\bar{\lambda}}(t)$ for all t . In addition, it is clear from the construction that $\text{law} X_t^{opt} = \mu_t$.

Next, we define some functions which will be useful in the proof of our main result. For each $t \in [0, 1]$, we define a function $g_t : \mathbb{R} \rightarrow \mathbb{R}$ by

$$g_t(x) = h'\left(\int_0^1 x_\lambda(r) dr\right),$$

where $\lambda = \mu_t((-\infty, x)) = F_t(x)$. We then set

$$G_t(x) = \int_{x_{\frac{1}{2}}(t)}^x g_t(y) dy$$

As h is convex, $x \mapsto h'(x)$ is increasing. As $x \mapsto F_t(x) = \lambda$ and $\lambda \mapsto \int_0^1 x_\lambda(t)dt$ are also increasing, it follows that $x \mapsto g_t(x)$ is an increasing function of x and so $x \mapsto G_t(x)$ is convex.

Finally, as the function $\lambda \mapsto \int_0^1 x_\lambda(t)dt$ is monotone, its (possibly infinite) limit as λ increases to 1 exists: denote $I_{max} = \lim_{\lambda \rightarrow 1^-} \int_0^1 x_\lambda(t)dt$. Similarly, define $I_{min} = \lim_{\lambda \rightarrow 0^+} \int_0^1 x_\lambda(t)dt$.

3 Main Result

3.1 Statement and proof of main result

Our main result is the following.

Theorem 3.1.1. *The process X_t^{opt} defined in equation (1) is the unique maximizer in (MK_∞) .*

The proof requires several Lemmas.

Lemma 3.1.2. *For any process $X_t \in \Pi$, $E(\int_0^1 G_t(X_t)dt) = \int_0^1 \int_{\mathbb{R}} G_t(x)d\mu_t(x)dt$*

Proof. The result is a simple application of Fubini's theorem:

$$\begin{aligned} E(\int_0^1 G_t(X_t)dt) &= \int_0^1 E(G_t(X_t))dt \\ &= \int_{\mathbb{R}} \int_0^1 G_t(x)d\mu_t(x)dt \end{aligned}$$

□

Lemma 3.1.3. *For each $\lambda \in (0, 1)$, the curve $x_\lambda(\cdot)$ maximizes the functional*

$$\mathcal{F}(\tilde{x}(\cdot)) = h(\int_0^1 \tilde{x}(t)dt) - \int_0^1 G_t(\tilde{x}(t))dt \quad (2)$$

among all continuous curves $\tilde{x}(\cdot)$ on $[0, 1]$, with $\tilde{x}(t) \in (a_t, b_t)$ for all t .

Proof. Our strategy will be to approximate these integrals by Riemann sums, and exploit known results on the corresponding optimal transportation problem with finitely many marginals. For a positive integer N , set $t_i = \frac{i}{N}$, for $i = 1, 2, \dots, N$ and consider the N -marginal optimal transportation problem with marginals μ_{t_i} and surplus function

$$b(x_{t_1}, x_{t_2}, \dots, x_{t_N}) = h(\frac{\sum_{i=1}^N x_{t_i}}{N}).$$

Note that this surplus function satisfies the strictly two monotone condition in [5] (essentially that $\frac{\partial^2 b}{\partial x_{t_i} \partial x_{t_j}} > 0$ for all $i \neq j$); by a result of Carlier [5] (extended to unbounded domains in [26]), the solution is concentrated on the set

$$\{(x_\lambda(t_1), x_\lambda(t_2), \dots, x_\lambda(t_N)) : \lambda \in (0, 1)\}.$$

It is well known that there exists a solution $(\phi_{t_1}^N, \phi_{t_2}^N, \dots, \phi_{t_N}^N)$ to the *Kantorovich dual problem* [17]. These functions satisfy:

$$\sum_{i=1}^N \phi_{t_i}^N(x_{t_i}) \geq h\left(\frac{\sum_{i=1}^N x_{t_i}}{N}\right)$$

with equality on the set $\{(x_\lambda(t_1), x_\lambda(t_2), \dots, x_\lambda(t_N)) : \lambda \in (0, 1)\}$.

Furthermore, due to the convexity of h , the functions $\phi_{t_i}^N$ are all convex, and due to assumptions 2 and 3, they are differentiable and satisfy the first order condition

$$\frac{1}{N} h'\left(\frac{\sum_{i=1}^N x_\lambda(t_i)}{N}\right) = \phi_{t_i}^N{}'(x_\lambda(t_i)),$$

for all $\lambda \in (0, 1)$.

Up to an additive constant, we can assume that each $\phi_{t_i}^N(x_{\frac{1}{2}}(t_i)) = 0$

Now, let $t \rightarrow \tilde{x}(t)$ be any continuous curve such that $\tilde{x}(t) \in (a_t, b_t)$ for all $t \in [0, 1]$. Then we have, for any λ :

$$h\left(\frac{\sum_{i=1}^N \tilde{x}(t_i)}{N}\right) - \sum_{i=1}^N \phi_{t_i}^N(\tilde{x}(t_i)) \leq h\left(\frac{\sum_{i=1}^N x_\lambda(t_i)}{N}\right) - \sum_{i=1}^N \phi_{t_i}^N(x_\lambda(t_i)) \quad (3)$$

For each $s \in [0, 1]$, set $\lambda_s = F_s(\tilde{x}(s))$. We then have $\lambda_s \in (0, 1)$ and $x_{\lambda_s}(s) = \tilde{x}(s)$. By assumptions 2 and 4, the map $s \mapsto \lambda_s$ is continuous. Therefore, there exists some $0 < c \leq d < 1$ such that $c \leq \lambda_s \leq d$ for all $0 \leq s \leq 1$.

Therefore, by the monotonicity of $\lambda \mapsto x_\lambda(t)$, we have $x_c(t) \leq x_{\lambda_s}(t) \leq x_d(t)$, and so, for all $s \in [0, 1]$,

$$I_c := \int_0^1 x_c(t) dt \leq \int_0^1 x_{\lambda_s}(t) dt \leq \int_0^1 x_d(t) dt =: I_d$$

Therefore, $s \mapsto \int_0^1 x_{\lambda_s}(t) dt$ is bounded above and below; choose K such that $|h'(x)| \leq K$ whenever $I_c - 1 \leq x \leq I_d + 1$

Now, for each point $x_{t_i} \in (a_{t_i}, b_{t_i})$ such that $\lambda := F_{t_i}(x_{t_i}) \in [c, d]$, we have

$$|\phi_{t_i}^N{}'(x_{t_i}) - \frac{1}{N} g_{t_i}(x_{t_i})| = \frac{1}{N} |h'\left(\frac{\sum_{j=1}^N x_\lambda(t_j)}{N}\right) - h'\left(\int_0^1 x_\lambda(t) dt\right)| \quad (4)$$

$$\leq K \frac{1}{N} \left| \frac{\sum_{j=1}^N x_\lambda(t_j)}{N} - \int_0^1 x_\lambda(t) dt \right| \quad (5)$$

Choose $\epsilon > 0$. By uniform convergence of the Riemann sums of $x_\lambda(t)$ (assumption 6), we can choose N so large that we have

$$\left| \frac{\sum_{i=1}^N x_\lambda(t_i)}{N} - \int_0^1 x_\lambda(t) dt \right| < \epsilon \quad (6)$$

for all $\lambda \in [c, d]$.

Now, the continuous function $t \mapsto \tilde{x}(t) - x_{\frac{1}{2}}(t)$ is bounded on $[0, 1]$:

$$|\tilde{x}(t) - x_{\frac{1}{2}}(t)| \leq M, \quad (7)$$

and so integrating $\phi_{t_i}^N(x_{t_i})$ and $g_{t_i}(x_{t_i})$ from $x_{\frac{1}{2}}(t_i)$ to $\tilde{x}(t_i)$ and combining inequalities (5), (6), and (7) yields

$$\begin{aligned} |\phi_{t_i}^N(\tilde{x}(t_i)) - \frac{1}{N}G_{t_i}(\tilde{x}(t_i))| &= \leq \int_{x_{\frac{1}{2}}(t_i)}^{\tilde{x}(t_i)} |\phi_{t_i}^{N'}(x) - \frac{1}{N}g_{t_i}(x)| dx \\ &\leq KM \frac{1}{N} \epsilon \end{aligned}$$

We then have

$$|\sum_{i=1}^N \phi_{t_i}^N(\tilde{x}(t_i)) - \sum_{i=1}^N \frac{1}{N}G_{t_i}(\tilde{x}(t_i))| \leq KM\epsilon$$

Now, the term $\sum_{i=1}^N \frac{1}{N}G_{t_i}(\tilde{x}(t_i))$ converges to $\int_0^1 G_t(\tilde{x}(t))dt$ as $N \rightarrow \infty$ and so the above inequality implies that $\sum_{i=1}^N \phi_{t_i}^N(\tilde{x}(t_i))$ converges to $\int_0^1 G_t(\tilde{x}(t))dt$ as well. Similarly, $\sum_{i=1}^N \phi_{t_i}^N(x_{\lambda}(t_i))$ converges to $\int_0^1 G_t(x_{\lambda}(t))dt$. As $h(\frac{\sum_{i=1}^N \tilde{x}(t_i)}{N})$ converges to $h(\int_0^1 \tilde{x}(t)dt)$ and $h(\sum_{i=1}^N x_{\lambda}(t_i))$ converges to $h(\int_0^1 x_{\lambda}(t)dt)$ the result now follows by taking the limit of (3). \square

Lemma 3.1.4. *For each $\lambda \in (0, 1)$, the curve $x_{\lambda}(\cdot)$ maximizes the functional*

$$\mathcal{F}(\tilde{x}(\cdot)) = h(\int_0^1 \tilde{x}(t)dt) - \int_0^1 G_t(\tilde{x}(t))dt \quad (8)$$

among all continuous curves $\tilde{x}(t)$ on $[0, 1]$, with $\tilde{x}(t) \in \text{spt}(\mu_t)$ for all t .

Proof. By the preceding lemma, we only need to consider the case where $\tilde{x}(t) = b_t$ or a_t for some t . We will assume here that $\tilde{x}(t) = b_t$ for some t , but $a_t < \tilde{x}(t)$ for all t (the proofs of the other cases are similar).

First, suppose that $\lim_{x \rightarrow I_{max}^+} h'(x) > 0$. This implies that $\lim_{x \rightarrow b_t} g_t(x) > 0$ for each t . Then, for $\epsilon > 0$, consider the set

$$A_{\epsilon} = \{t : \tilde{x}(t) - \epsilon \in (a_t, b_t)\} \cap \{t : g_t(\tilde{x}(t) - \epsilon) > 0\}$$

As an open subset of $[0, 1]$, A_{ϵ} can be written as a countable union of disjoint intervals, $A_{\epsilon} = \cup_i I_i$. For each i , let $q_i(t) : I_i \rightarrow \mathbb{R}$ be a continuous function such that $q_i(t) = 0$ on the boundary of I_i and $0 < q_i(t) < 1$ on the interior of I_i . We define the function

$$\begin{aligned} \tilde{x}^{\epsilon}(t) &= \tilde{x}(t) - \epsilon q_i(t), \text{ if } t \in I_i \text{ for some } i \\ \tilde{x}^{\epsilon}(t) &= \tilde{x}(t), \text{ otherwise.} \end{aligned}$$

Note that $\tilde{x}^\epsilon(t) \leq \tilde{x}(t)$ everywhere, and the inequality is strict for $t \in A_\epsilon$. As $x \mapsto g_t(x)$ is monotone increasing and $g_t(\tilde{x}(t) - \epsilon) > 0$ on A_ϵ , we have $g_t(x) > 0$ for $\tilde{x}(t) - \epsilon < x < \tilde{x}(t)$. Thus, $\frac{dG_t}{dx}(x) = g_t(x) > 0$, which implies $G_t(\tilde{x}^\epsilon(t))$ increases as ϵ decreases. Furthermore, $|\tilde{x}^\epsilon(t) - \tilde{x}(t)| \leq \epsilon$, so $\tilde{x}^\epsilon(t) \rightarrow \tilde{x}(t)$ as $\epsilon \rightarrow 0$. By continuity, $G_t(\tilde{x}^\epsilon(t)) \rightarrow G_t(\tilde{x}(t))$.

As the points where $\tilde{x}_t = b_t$ all lie in A_ϵ , we have $a_t < \tilde{x}^\epsilon(t) < b_t$ everywhere. Therefore, Lemma 3.1.3

$$h\left(\int_0^1 \tilde{x}^\epsilon(t) dt\right) - \int_0^1 G_t(\tilde{x}^\epsilon(t)) dt \leq h\left(\int_0^1 x_\lambda(t) dt\right) - \int_0^1 G_t(x_\lambda(t)) dt$$

Taking the limit as $\epsilon \rightarrow 0$, using the monotone convergence theorem on $\int_0^1 \tilde{x}^\epsilon(t) dt$ and $\int_0^1 G_t(\tilde{x}^\epsilon(t)) dt$ and the continuity of h yields the desired result.

Finally, if $\lim_{x \rightarrow I_{max}^+} h'(x) \leq 0$, then $g_t(x) < 0$ for all $x \in (a_t, b_t)$. A similar argument to the above implies that $\epsilon \mapsto G_t(\tilde{x}^\epsilon(t))$ as ϵ decreases and applying the monotone convergence theorem again yields the desired result. \square

Lemma 3.1.5. *The curves $x_\lambda(t)$ are the only maximizers of the functional $\mathcal{F}(\tilde{x}(\cdot))$ in Lemma 3.1.4.*

Proof. To show that no other continuous curve \tilde{x} can be a maximizer, we use a variational argument. Assume that $\tilde{x}(\cdot)$ is a curve such that $\tilde{x}(t) \in \text{spt}(\mu_t)$ for all t , maximizing (8). Let $y(t)$ be a continuous function such that $\tilde{x}(t) + \epsilon y(t) \in \text{spt}(\mu_t)$ for sufficiently small $\epsilon > 0$. We then have:

$$\begin{aligned} 0 > \frac{d\mathcal{F}(\tilde{x}(t) + \epsilon y(t))}{d\epsilon^+} \Big|_{\epsilon=0} &= \frac{d}{d\epsilon^+} \Big|_{\epsilon=0} \left[h\left(\int_0^1 \tilde{x}(t) + \epsilon y(t) dt\right) - \int_0^1 G_t(\tilde{x}(t) + \epsilon y(t)) dt \right] \\ &= h'\left(\int_0^1 \tilde{x}(t) dt\right) \int_0^1 y(t) dt - \int_0^1 g_t(\tilde{x}(t)) y(t) dt \\ &= \int_0^1 \left[h'\left(\int_0^1 \tilde{x}(r) dr\right) - g_t(\tilde{x}(t)) \right] y(t) dt \end{aligned} \quad (9)$$

Note that we have used Lebesgue's dominated convergence theorem on the last term.

Now, choose λ such that $\int_0^1 x_\lambda(s) ds = \int_0^1 \tilde{x}(s) ds$; we will show that $x_\lambda(t) = \tilde{x}(t)$ for all t . The proof is by contradiction; assume $x_\lambda(t) \neq \tilde{x}(t)$ for some fixed t . Suppose $x_\lambda(t) > \tilde{x}(t)$ (the case $x_\lambda(t) < \tilde{x}(t)$ is similar and is omitted). Then by continuity, $x_\lambda(s) > \tilde{x}(s)$ in some neighbourhood U of t . By monotonicity of $x \mapsto g_s(x)$, this implies that $g_s(x_\lambda(s)) > g_s(\tilde{x}(s))$ in U .

Now, as $\tilde{x}(s) < x_\lambda(s) \leq b_s$ on U , we may choose $y(s) > 0$ on U and $y(s) = 0$

elsewhere and have $\tilde{x}(s) + \epsilon y(s) \in \text{spt}(\mu_s)$ everywhere. Then we have, for $s \in U$

$$\begin{aligned}
\left[h' \left(\int_0^1 \tilde{x}(r) dr \right) - g_s(\tilde{x}(s)) \right] y(s) &= \left[h' \left(\int_0^1 x_\lambda(r) dr \right) - g_s(\tilde{x}(s)) \right] y(s) \\
&> \left[h' \left(\int_0^1 x_\lambda(r) dr \right) - g_s(x_\lambda(s)) \right] y(s) \\
&= \left[h' \left(\int_0^1 x_\lambda(r) dr \right) - h' \left(\int_0^1 x_\lambda(r) dr \right) \right] y(s) \\
&= 0
\end{aligned}$$

As $y(s) = 0$ outside of U , this yields

$$\int_0^1 \left[h' \left(\int_0^1 \tilde{x}(r) dr \right) - g_s(\tilde{x}(s)) \right] y(s) ds > 0$$

This contradicts inequality (9), completing the proof. \square

We are now in a position to prove Theorem 3.1.1.

Proof. It is clear from the definition that $X_t^{opt} \in \Pi$; that is, that $\text{law} X_t^{opt} = \mu_t$. Now, by Lemma 3.1.4, given any process $Y_t \in \Pi$ with single time marginals μ_t , we have, for each individual sample path and each λ ,

$$\begin{aligned}
h \left(\int_0^1 Y_t dt \right) &\leq h \left(\int_0^1 x_\lambda(t) dt \right) - \int_0^1 G_t(x_\lambda(t)) dt + \int_0^1 G_t(Y_t) dt \\
&= M + \int_0^1 G_t(Y_t) dt
\end{aligned}$$

where M denotes the minimal value of the functional \mathcal{F} in Lemma 3.1.4. Note that we have equality if and only if the sample path $Y_t = x_\lambda(t)$ for some λ , by Lemma 3.1.5. Taking expectations and using Lemma 3.1.2 yields

$$\begin{aligned}
E \left(h \left(\int_0^1 Y_t dt \right) \right) &\leq M + E \left(\int_0^1 G_t(Y_t) dt \right) \\
&= M + \int_0^1 \int_{\mathbb{R}} G_t(x) d\mu_t(x) dt.
\end{aligned}$$

and we have equality if and only if $Y_t = x_\lambda(t)$ for some λ , almost surely. Now, if we interpret a stochastic process as a measure on the set of continuous functions, it is clear that only one measure with marginals μ_t can be supported on the set $\{x_\lambda(\cdot) : \lambda \in [0, 1]\}$; namely, the optimizer X_t^{opt} . This completes the proof. \square

3.2 Extensions of the main theorem

In this subsection we present two extensions of our main theorem, designed specifically with potential applications in mind.

The first extension is a relaxation of the smoothness and strict convexity of h . One motivation for this result is to find an upper bound on the price of an Asian option, whose value is given by $E[h(\int_0^1 X_t dt)]$, for a particular process X_t with $h(x) = \max(0, ax - b)$ (ie, a non smooth, convex but not strictly convex surplus function). In this case, we potentially lose uniqueness in our main theorem.

Theorem 3.2.1. *Suppose that h is convex. Assume conditions 2-6, and in addition assume that the second moments of μ_t exist and are integrable, that is, $\int_0^1 \int_{\mathbb{R}} x^2 d\mu_t(x) dt < \infty$. Then the process X_t^{opt} defined in section 2 is optimal for (MK_∞) .*

Proof. Let h_n be a sequence of smooth, strictly convex functions such that $h_n(x) \geq h_{n+1}(x)$, $h_n(x) \rightarrow h(x)$ pointwise and

$$h(x) \leq h_n(x) \leq h(x) + 1 + x^2,$$

Then h_n satisfies assumptions 1 and 5, and therefore, for any $Y_t \in \Pi$ we have, by Theorem 3.1.1

$$E\left(h_n\left(\int_0^1 Y_t dt\right)\right) \leq E\left(h_n\left(\int_0^1 X_t^{opt} dt\right)\right)$$

Now, by the monotone convergence theorem, $E\left(h_n\left(\int_0^1 Y_t dt\right)\right)$ converges to $E\left(h\left(\int_0^1 Y_t dt\right)\right)$, and similarly $E\left(h_n\left(\int_0^1 X_t^{opt} dt\right)\right)$ converges to $E\left(h\left(\int_0^1 X_t^{opt} dt\right)\right)$, implying the desired result. \square

Our second extension concerns coupling a stochastic process X_t with prescribed single time marginals and a real valued random variable Z with a prescribed law, to maximize $E[h(\int_0^1 X_t dt)Z]$. The motivation here is the Feynman-Kac formula; expressions of this form represent solutions to parabolic PDE and in the next section, we use our extension here to prove upper bounds on these solutions. Our extension here holds when Z is positive almost surely. If Z is merely non-negative almost surely, the process we construct is still optimal, but we possibly lose uniqueness.

Let μ_t be a curve of measures, satisfying the conditions in the second section. Let ν be a measure on the upper half line $\mathbb{R}_+ = \{y > 0\}$, absolutely continuous with respect to Lebesgue measure. Then consider the problem of finding a pair (X_t, Z) , where X_t is a continuous stochastic process, with law $X_t = \mu_t$ and Z a random variable with law $Z = \nu$, maximizing

$$E\left(h\left(\int_0^1 X_t dt\right)Z\right) \quad (MK_{\infty+1})$$

Recall the stochastic process X_t^{opt} from the section 2 and define the random variable $Z^{opt} = Z^{opt}(\lambda)$ on the same underlying probability space $[0, 1]$ implicitly by

$$\nu(-\infty, Z^{opt}(\lambda)) = \lambda$$

Then the pair (X_t^{opt}, Z^{opt}) satisfies $\text{law} X_t^{opt} = \mu_t$ and $\text{law} Z^{opt} = \nu$. We also define

$$\bar{g}(z) = h'(\int_0^t x_\lambda(t) dt) z_\lambda,$$

where $\lambda = \nu((-\infty, z_\lambda))$. We then set

$$\bar{G}(z) = \int_{z^{1/2}}^z g_t(y) dy$$

Theorem 3.2.2. *Assume conditions 1-6 from section 2 and that the measure ν on $(0, \infty)$ is absolutely continuous with respect to Lebesgue measure. Then the pair (X_t^{opt}, Z^{opt}) is the unique optimizer for $(MK_{\infty+1})$.*

Proof. The proof is similar to the proof of the main theorem, and is only sketched here. The key adjustment is in Lemma 3.1.3, where we replace $h(\sum_{i=1}^N x_{t_i})$ with $h(\sum_{i=1}^N x_{t_i})z$, where $z > 0$. This surplus function is again strictly two monotone, and so the result of Carlier [5] applies to the $N+1$ marginal optimal transportation problem with this surplus function and marginals $\mu_{t_1}, \mu_{t_2}, \dots, \mu_{t_N}, \nu$. Letting $\phi_{t_1}^N, \phi_{t_2}^N, \dots, \phi_{t_N}^N, u^N$ solve the dual problem, we obtain, for any curve $\tilde{x}(t)$, point $\tilde{z} > 0$ and $\lambda \in (0, 1)$:

$$h(\frac{\sum_{i=1}^N \tilde{x}(t_i)}{N})\tilde{z} - \sum_{i=1}^N \phi_{t_i}^N(\tilde{x}(t_i)) - u^N(\tilde{z}) \leq h(\frac{\sum_{i=1}^N x_\lambda(t_i)}{N})z_\lambda - \sum_{i=1}^N \phi_{t_i}^N(x_\lambda(t_i)) - u^N(z_\lambda).$$

From the first order conditions arising in the $N+1$ marginal problem, we have $u^{N'}(z) = h'(\frac{\sum_{i=1}^N x_\lambda(t_i)}{N})z_\lambda$, where $\lambda = \nu((-\infty, z_\lambda))$. Much like in the proof of Lemma 3.1.3, one can use this to show that as $N \rightarrow \infty$, $u^N(\tilde{Z}) \rightarrow \bar{G}(\tilde{Z})$. As in the proof of Lemma 3.1.3, we can take the limit as $N \rightarrow \infty$ and obtain

$$h(\int_0^1 \tilde{x}(t) dt)\tilde{y} - \int_0^1 \bar{G}_t(\tilde{x}(t)) dt - u(\tilde{y}) \leq h(\int_0^1 x_\lambda(t) dt)\tilde{y} - \int_0^1 \bar{G}_t(x_\lambda(t)) dt - u(y_\lambda).$$

Here, $\bar{G}_t := \int_{x_{\frac{1}{2}}(t)}^x g_t(y) dy$ is defined as in section 2, except that we take $\bar{g}_t(x) = h'(\int_0^1 x_\lambda(r) dr)z_\lambda$, for $\lambda = F_t(x)$, whereas in section 2 we took $g(x) = h'(\int_0^1 x_\lambda(r) dr)$.

The remainder of the proof is analogous to the proof of the main theorem and preceding lemmas in subsection 3.1 and is omitted. \square

4 Rearrangement inequalities and applications

4.1 A rearrangement inequality

Carlier observed that a solution to the multi-marginal problem in 1-dimension yields a refinement of the Hardy-Littlewood inequality [5]. Here, we use our main theorem to develop an analogous infinite dimensional rearrangement inequality of Hardy-Littlewood type. Given a stochastic process Y_t , set $\mu_t = \text{law}(Y_t)$. We can then find the process X_t^{opt} corresponding to these marginals, as in section 2 (that is, a monotone rearrangement of Y_t) and, assuming the marginals and convex h satisfy assumptions 1-6, Theorems 3.1.1 and 3.2.2 yield the following result, which can be viewed as an infinite dimensional version of the Hardy-Littlewood inequality

Proposition 4.1.1. *Let Y_t be a stochastic process whose marginals $\mu_t = \text{law}(Y_t)$ satisfy conditions 1-6. Let Z be a random variable on \mathbb{R}^+ and assume $\nu = \text{law}(Z)$ is absolutely continuous with respect to Lebesgue. Let $(X_t^{\text{opt}}, Z^{\text{opt}})$ be the optimal process corresponding to these marginals, defined in section 2 (X_t^{opt}) and subsection 3.2 (Z^{opt}). We then have*

$$\begin{aligned} E(h(\int_0^1 Y_t dt)) &\leq E(h(\int_0^1 X_t^{\text{opt}} dt)) \\ &= \int_0^1 h(\int_0^1 x_\lambda(t) dt) d\lambda \end{aligned}$$

and

$$\begin{aligned} E(h(\int_0^1 Y_t dt)Z) &\leq E(h(\int_0^1 X_t^{\text{opt}} dt)Z^{\text{opt}}) \\ &= \int_0^1 h(\int_0^1 x_\lambda(t) dt) z_\lambda d\lambda \end{aligned}$$

It is interesting to note that Jensen's inequality also yields upper bounds on quantities of the form $E(h(\int_0^1 Y_t dt))$, and these upper bounds are often used in finance and physics (see the examples in the subsequent section). Recall that for each sample path Y_t of a stochastic process and convex h , Jensen's inequality says:

$$h(\int_0^1 Y_t dt) \leq \int_0^1 h(Y_t) dt$$

so that, for any stochastic process Y_t , we have:

$$E(h(\int_0^1 Y_t dt)) \leq E(\int_0^1 h(Y_t) dt)$$

We now prove our bound (4.1.1) is *tighter* than the bound coming from Jensen's inequality.

Proposition 4.1.2. *Given a stochastic process Y_t , let X_t^{opt} be the optimal process from section 2, with marginals $\mu_t = \text{law}(Y_t)$. Then:*

$$E(h(\int_0^1 X_t^{opt} dt)) \leq E(\int_0^1 h(Y_t) dt)$$

Proof. We apply Jensen's inequality to $E(h(\int_0^1 X_t^{opt} dt))$, and use Fubini's theorem on the resulting expression, yielding

$$\begin{aligned} E(h(\int_0^1 X_t^{opt} dt)) &\leq E(\int_0^1 h(X_t^{opt} dt)) \\ &= \int_0^1 E(h(X_t^{opt})) dt \\ &= \int_0^1 \int_{\mathbb{R}} h(x) d\mu_t(x) dt \\ &= \int_0^1 \int_{\mathbb{R}} h(y) d\mu_t(y) dt \\ &= \int_0^1 E(h(Y_t)) dt \\ &= E(\int_0^1 h(Y_t) dt) \end{aligned}$$

where we have used the fact that the $\text{law}(Y_t) = \text{law}(X_t^{opt}) = \mu_t$. \square

Also of note is that, by applying Jensen's inequality to the expectation E rather than the integral \int_0^1 , we obtain a *lower* bound on $E(h(\int_0^1 Y_t))$, which by Fubini's theorem, depends only on the marginals $\text{law}(Y_t)$. Combining our main theorem with Jensen's inequality in this way, then, gives upper and lower bounds for $E(h(\int_0^1 Y_t))$ which any process Y_t with marginals μ_t must satisfy.

4.2 Example

We illustrate the rearrangement inequality in Proposition 4.1.1 with an example. In this example, the convex function $h(x) = e^x$ is the exponential function.

Proposition 4.2.1. *Let W_t be Brownian motion starting at $W_0 = 0$ and $\mu_t = \text{law}(W_t)$, so that $d\mu_t(x) = \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} dx$. Then, for $s > 0$,*

$$\begin{aligned} E(e^{\int_0^s W_t dt}) &= e^{\frac{s^3}{6}} \\ E(e^{\int_0^s X_t^{opt} dt}) &= e^{\frac{2s^3}{9}} \\ E(\frac{1}{s} \int_0^s e^{sW_t} dt) &= \frac{2}{s^3} [e^{\frac{s^3}{2}} - 1] \end{aligned}$$

After rescaling the interval $[0, 1]$ to $[0, s]$, these three quantities represent, respectively, $E(e^{\int_0^s W_t dt})$, the upper bound from Proposition 4.1.1 (see the Remark below), and the upper bound from Jensen's inequality, respectively. It is straightforward to verify directly that $e^{\frac{s^3}{6}} \leq e^{\frac{2s^3}{9}} \leq \frac{2}{s^3}[e^{\frac{s^3}{2}} - 1]$, as is implied by Propositions 4.1.1 and 4.1.2.

Remark 4.2.2. *Strictly speaking, our main theorem does not apply here, as the marginal μ_0 is a Dirac mass at $x = 0$ and not absolutely continuous with respect to Lebesgue (the other conditions from section 2 are easy to verify). However, Lemma 3.1.3 implies that for any $\epsilon > 0$ and any curve $\tilde{x}(t)$ we have*

$$h(\int_{\epsilon}^1 \tilde{x}_t dt) - \int_{\epsilon}^1 g_t(\tilde{x}_t) dt \leq h(\int_{\epsilon}^1 x_{\lambda}(t) dt) - \int_{\epsilon}^1 g_t(x_{\lambda}(t)) dt$$

Taking the limit as $\epsilon \rightarrow 0$ yields:

$$h(\int_0^1 \tilde{x}_t dt) - \int_0^1 g_t(\tilde{x}_t) dt \leq h(\int_0^1 x_{\lambda}(t) dt) - \int_0^1 g_t(x_{\lambda}(t)) dt$$

We can then prove, just as in Section 3, that for any process $Y_t \in \Pi$, we have

$$E(h(\int_0^1 Y_t dt)) \leq E(h(\int_0^1 X_t^{opt} dt))$$

Proof. Let h be the exponential function. The curves $x_{\lambda}(t)$ here satisfy

$$\lambda = \int_{-\infty}^{x_{\lambda}(t)} \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy = \int_{-\infty}^{\frac{x_{\lambda}(t)}{\sqrt{t}}} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$$

which implies $x_{\lambda}(t) = \sqrt{t}z_{\lambda}$, where $\lambda = \text{erf}(z_{\lambda})$. Therefore, $\int_0^s x_{\lambda}(t) dt = \frac{2}{3}z_{\lambda}s^{\frac{3}{2}}$. Our expectation becomes:

$$\begin{aligned} E(h(\int_0^s X_t^{opt} dt)) &= \int_0^1 e^{\frac{2}{3}z_{\lambda}s^{\frac{3}{2}}} d\lambda \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{2}{3}zs^{\frac{3}{2}}} e^{-\frac{z^2}{2}} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(z - \frac{2s^{\frac{3}{2}}}{3})^2 - 4\frac{s^3}{9}]} dz \\ &= \frac{e^{\frac{2s^3}{9}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z - \frac{4s^{\frac{3}{2}}}{3})^2} dz \\ &= e^{\frac{2s^3}{9}} \end{aligned}$$

Now, as W_t is Brownian motion, it is well known that $Z_s = \int_0^s W_t dt$ is normally distributed, with mean zero and variance $\frac{s^3}{3}$. Therefore, we have that

$$\begin{aligned}
E(e^{\int_0^s W_t dt}) &= \sqrt{\frac{3}{2\pi s^3}} \int_{-\infty}^{\infty} e^z e^{-\frac{3z^2}{2s^3}} dz \\
&= e^{\frac{s^3}{6}} \sqrt{\frac{3}{2\pi s^3}} \int_{-\infty}^{\infty} e^{-\frac{3(z-\frac{s^3}{3})^2}{2s^3}} dz \\
&= e^{\frac{s^3}{6}}
\end{aligned}$$

Finally, we compute the upper bound from Jensen's inequality.

$$\begin{aligned}
E(e^{\int_0^s W_t dt}) &\leq E\left(\frac{1}{s} \int_0^s e^{sW_t} dt\right) \\
&= \frac{1}{s} \int_0^s E(e^{sW_t}) dt \\
&= \frac{1}{s} \int_0^s \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{sW} e^{-\frac{W^2}{2t}} dW dt \\
&= \frac{1}{s} \int_0^s e^{\frac{s^2 t}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(W-st)^2}{2t}} dW dt \\
&= \frac{1}{s} \int_0^s e^{\frac{s^2 t}{2}} ds \\
&= \frac{2}{s^3} [e^{\frac{s^3}{2}} - 1]
\end{aligned}$$

□

4.3 Upper bounds on solutions to Parabolic PDE

Consider the partial differential equation on $(\vec{x}, t) \in \mathbb{R}^{n+1}$:

$$\frac{\partial f}{\partial t} + \sum_{i=1}^n b_i(\vec{x}, t) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (A^T A)_{ij}(\vec{x}, t) \frac{\partial^2 f}{\partial x_i \partial x_j} = V(\vec{x}, t) f$$

with terminal condition $f(\vec{x}, T) = \psi(\vec{x})$. Here, $b = b(\vec{x}, t)$ is an \mathbb{R}^n valued function and $A = A(\vec{x}, t)$ an $n \times n$ matrix valued function. Equations of this type have numerous applications in mathematical finance, mathematical physics, biology, etc. According to the Feynman-Kac formula, the solution is given by:

$$f(\vec{x}, s) = E[e^{-\int_s^T V(\vec{X}_t) dt} \psi(\vec{X}_T)] \quad (10)$$

where \vec{X}_t is the solution of the stochastic PDE

$$d\vec{X} = b(\vec{X}, t) dt + A(\vec{X}, t) d\vec{W}, \quad \vec{X}_s = \vec{x}$$

Here \vec{W} is a Wiener process on \mathbb{R}^n . Suppose that $\psi > 0$. Letting $\mu_t = \text{law}(V(\vec{X}_t))$ and $\nu = \text{law}(\psi(\vec{X}_T))$, then, if assumptions 1-6 are satisfied, Proposition 4.1.1 tells us that:

$$f(\vec{x}, s) \leq E[e^{-\int_s^T X_t^{opt} dt} Z^{opt}] \quad (11)$$

$$= \int_0^1 e^{-\int_s^T x_\lambda(t) dt} z_\lambda d\lambda \quad (12)$$

Depending on b , A and V , the Feynman-Kac formula may be very difficult to evaluate analytically. In practice, the integral (10) is often computed numerically using Monte-Carlo techniques. In certain cases, however, the single time marginals of the real-valued process $V(\vec{X}_t)$ can be determined. To find an upper bound on $f(\vec{x}, t)$, then, we must only compute the integral (11), which boils down to computing two, 1-dimensional integral. Even if this integral cannot easily be calculated analytically, it is a much simpler task to compute numerically than the *infinite* dimensional integral in (10). In the next subsection, we present a particular application in quantum physics.

5 Application: phase space bounds in quantum physics

In this subsection, we illustrate how our result can be used to prove refinements of two phase-space bounds in quantum physics. A discussion of the physical significance of these bounds can be found in the book of Simon [3]. The basic idea behind their proofs is that, as outlined by Simon, phase space bounds can be derived by applying Jensen's inequality to quantities of the form (MK_∞) . Our general strategy here is to mimic these proofs, replacing Jensen's inequality with the result from Proposition 4.1.1; as a consequence of Proposition 4.1.2, our phase space bounds here are tighter.

Let V be a continuous, integrable, function on \mathbb{R}^n which is bounded from below. For fixed t , let W_s^t be a Brownian bridge process in s on \mathbb{R}^n , starting and ending at $\vec{x} = 0$, at times $s = 0$ and $s = t$, respectively; that is, $W_0^t = 0$ and $W_t^t = 0$.

Let h be the exponential function and, for each $\vec{x} \in \mathbb{R}^n$, assume that the measures $\mu_{s,t,\vec{x},V} = \text{law}[V(\vec{x} + W_s^t)]$ satisfy conditions 1-6¹. Let $Y_{s,t,\vec{x},V}^\lambda$ be defined by the implicit relationship

$$\lambda = \mu_{s,t,\vec{x},V}(-\infty, Y_{s,t,\vec{x},V}^\lambda).$$

Let H_0 be $-\frac{1}{2}\Delta$, where Δ is the Laplacian.

The following classical bound on the trace of the operator $e^{-t(H_0+V)}$ is due to Golden [14], Thompson [33] and Symanzik [32]:

$$\text{Tr}(e^{-t(H_0+V)}) \leq \int_{\mathbb{R}^n} \frac{1}{(2\pi t)^{n/2}} e^{-tV(\vec{x})} d\vec{x}, \quad (13)$$

¹Note that, strictly speaking, the measures $\mu_{s,t,\vec{x},V}$ cannot satisfy assumption 2, as $\mu_{0,t,\vec{x},V}$ and $\mu_{t,t,\vec{x},V}$ are Dirac masses at $V(\vec{x})$; however, this can be dealt with as in the example in subsection 4.4.

The proof of our bound below is based on Symanzik's proof of inequality (13), outlined in the book by Simon.

Theorem 5.0.1.

$$\text{Tr}(e^{-t(H_0+V)}) \leq \int_{\mathbb{R}^n} \int_0^1 e^{\int_0^t Y_{s,t,\vec{x},V}^\lambda ds} d\lambda d\vec{x},$$

Proof. We follow the argument in Simon, using our inequality from Theorem 4.1.1 in place of Jensen's inequality.

As in Simon, the Feynman-Kac formula implies:

$$\text{Tr}(e^{-t(H_0+V)}) = \int_{\mathbb{R}^3} E\left[e^{-\int_0^t V(\vec{x}+W_s^t)ds}\right] d\vec{x}.$$

Applying Proposition 4.1.1 to $E\left[e^{-\int_0^t V(\vec{x}+W_s^t)ds}\right]$ yields the desired result. \square

A similar argument yields a bound on the number of bound states, that is, the number $N(V)$ of negative eigenvalues of H_0 . Using Jensen's inequality, this leads to Lieb's proof of the celebrated Cwikel-Lieb-Rosenbljum bound [21], which states that, for $n \geq 3$, $N(V)$ is bounded from above by the $L^{n/2}$ norm of the negative part of V :

$$N(V) \leq a \int_{\mathbb{R}^n} |V_-(\vec{x})|^{n/2} d\vec{x}. \quad (14)$$

Let us mention that the independent proofs of Cwikel [8] and Rosenbljum [29] rely on different techniques. Other proofs of the CLR bound have since been found by Li and Yau [20], Fefferman [9] and Conlon [7].

Applying our main theorem instead leads to the bound below.

Theorem 5.0.2. *For any convex function h satisfying 1-6,*

$$N(V) \leq C(h) \int_0^\infty t^{-1} \int_{\mathbb{R}^n} \int_0^1 h\left(\int_0^t Y_{s,t,\vec{x},V}^\lambda ds\right) d\lambda d\vec{x} dt.$$

Here, $C(h)$ is a constant depending on h .

Proof. The argument of Lieb (again, reproduced by Simon) yields, for any convex function h on \mathbb{R}

$$N(V) \leq C(h) \int_0^\infty t^{-1} \int_{\mathbb{R}^n} E\left(h\left(\int_0^t V(W_s^t)ds\right)\right) d\vec{x} dt$$

We apply our inequality to $E\left(h\left(\int_0^t V(W_s^t)ds\right)\right)$, yielding the desired result. \square

Remark 1. *Let us note that the above holds for any convex function h . Lieb's proof goes on to apply Jensen's inequality to $h\left(\int_0^t V(W_s^t)ds\right)$ and obtains (14) with*

$$a = C(h) \frac{1}{2\pi} \int_0^\infty s^{-1-n/2} h(s) ds$$

If $n \geq 3$, it is possible to choose the function h so that both $C(h)$ and the preceding integral are finite [21][3]. In principal, one can then optimize over convex functions h to obtain as good a constant a as possible (note, however, that the optimal constant is not known). Note, however, that for whichever function h one chooses, our bound above, using the same h , is tighter.

It should be emphasized that, while our bounds appear complicated, they can be calculated for any convex function h simply by determining the single time marginals of $V[\vec{x} + W_s^t]$ and calculating several (two in Theorem 5.0.1 and three in Theorem 5.0.2) 1-dimensional integrals and one n -dimensional integral, which is relatively easy to do numerically even if the form of V makes it difficult analytically. On the other hand, our bounds *improve* the classical phase space bounds, as our inequality from Proposition 4.1.1 is always tighter than Jensen's inequality, by Proposition 4.1.2.

References

- [1] M. Beiglbock, P. Henry-Labordere, and F. Penkner. Model independent bounds for option prices: a mass transport approach. Preprint.
- [2] Y. Brenier. Decomposition polaire et rearrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris. Ser. I Math.*, 305:805–808, 1987.
- [3] B.Simon. *Functional integration and quantum physics*. Academic Press, New York, 1979.
- [4] L. Caffarelli. Allocation maps with general cost functions. In *Partial Differential Equations and Applications*, volume 177 of *Lecture Notes in Pure and Applied Math*, pages 29–35. Dekker, New York, 1996.
- [5] G. Carlier. On a class of multidimensional optimal transportation problems. *J. Convex Anal.*, 10(2):517–529, 2003.
- [6] G. Carlier and B. Nazaret. Optimal transportation for the determinant. *ESAIM Control Optim. Calc. Var.*, 14(4):678–698, 2008.
- [7] J.G. Conlon. A new proof of the Cwikel-Lieb-Rosenbljum bound. *Rocky Mountain J. Math*, 15:117–122, 1985.
- [8] M. Cwikel. Weak type estimates and the number of bound states of Schrodinger operators. *Ann. of Math.*, 106:93–102, 1977.
- [9] C. Fefferman. The uncertainty principle. *Bull. Amer. Math. Soc.*, 9:129–206, 1983.
- [10] A. Galichon, P. Henry-Labordere, and N. Touzi. A stochastic control approach to non-arbitrage bounds given marginals, with an application to Lookback options. Preprint available at <https://sites.google.com/site/alfredgalichon/research>.

- [11] W. Gangbo. PhD thesis, Universite de Metz, 1995.
- [12] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.
- [13] W. Gangbo and A. Święch. Optimal maps for the multidimensional monge-kantorovich problem. *Comm. Pure Appl. Math.*, 51(1):23–45, 1998.
- [14] S. Golden. Lower bounds for the Helmholtz function. *Phys. Rev. B*, 137:1127–1128, 1965.
- [15] F.R. Harvey and B. Lawson. Split Special-Lagrangian geometry. Preprint available at arxiv:1007.0450.
- [16] H. Heinich. Probleme de Monge pour n probabilities. *C.R. Math. Acad. Sci. Paris*, 334(9):793–795, 2002.
- [17] H.G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67:399–432, 1984.
- [18] M. Knott and C. Smith. On a generalization of cyclic monotonicity and distances among random vectors. *Linear Algebra Appl.*, 199:363–371, 1994.
- [19] V. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-Valued Analysis*, 7(1):7–32, 1999.
- [20] P. Li and S.T. Yau. On the Schrodinger equation and the eigenvalue problem. *Comm. Math. Phys.*, 88:309–318, 1983.
- [21] E. Lieb. Bounds on the eigenvalues of the Laplace and Schrodinger operators. *Bull. Amer. Math. Soc*, 82:751–753, 1976.
- [22] E. H. Lieb and W. E. Thirring. Bound for the kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.*, 35:687–689, Sep 1975.
- [23] R.J. McCann. A convexity theory for interacting gases and equilibrium crystals. *Adv. Math.*, 128(1):153–179, 1997.
- [24] I. Olkin and S.T. Rachev. Maximum submatrix traces for positive definite matrices. *SIAM J. Matrix Ana. Appl.*, 14:390–39, 1993.
- [25] B. Pass. Optimal transportation with infinitely many marginals. Preprint.
- [26] B. Pass. PhD thesis, University of Toronto, 2011.
- [27] B. Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

- [28] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 43:529–536, 2012. 10.1007/s00526-011-0421-z.
- [29] G. Rosenbljum. The distribution of the discrete spectrum for singular differential operators. *Dokl. Akad. Nauk SSSR*, 202:245–249, 1972.
- [30] L. Rüschendorf and L. Uckelmann. On optimal multivariate couplings. In *Proceedings of Prague 1996 conference on marginal problems*, pages 261–274. Kluwer Acad. Publ., 1997.
- [31] L. Rüschendorf and L. Uckelmann. On the n -coupling problem. *J. Multivariate Anal.*, 81:242–258, 2002.
- [32] K. Symanzik. Proof and refinement of an inequality of Feynman. *J. Math. Phys.*, 6:1155–1156, 1965.
- [33] C.T. Thompson. Inequality with applications in statistical mechanics. *J. Math. Phys.*, 6:1812–1813, 1965.
- [34] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.
- [35] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 2009.